R. Dutter · P. Filzmoser
U. Gather · P. J. Rousseeuw
Editors

Developments in Robust Statistics

International Conference on Robust Statistics 2001

With 85 Figures
and 75 Tables

Physica-Verlag
A Springer-Verlag Company

Robustness Analysis in Forecasting of Time Series

Y. Kharin

Belarussian State University, Minsk, Belarus
E-mail: kharin@bsu.by

Summary. The problems of statistical forecasting of time series under distortions of traditional hypothetical models are considered. The following distorted models of time series are used: trend models under "outliers" and functional distortions, regression models under "outliers" and "errors-in-regressors", autoregressive time series with parameter specification errors and non-homogeneous innovations. Robustness characteristics based on the mean square risk of forecasting are introduced and evaluated for these cases. In addition, new robust forecasting procedures are presented.

Key words: Robustness, Forecasting, Time series, Distortions, Risk

1 Introduction

Statistical forecasting procedures are intensively used to solve many significant problems in engineering, economics, finance, medicine, ecology, etc. (Abraham and Ledolter, 1989; Bowerman, 1993; Peña et al., 2001). Optimality or asymptotic optimality with increase in observation time of the majority of the known statistical forecasting procedures is proved w.r.t. the mean square risk of prediction under the assumptions of an underlying hypothetical model (Anderson, 1971; Hamilton, 1994).

In practice, however, the observed time series do not follow the hypothetical models (Hampel et al., 1986; Huber, 1981): random errors are non-Gaussian, correlated, non-homogeneous (Boutahar and Deniau, 1995; Dahlhaus and Wefelmeyer, 1996); data is corrupted by "outliers", "level shifts" or "missings" (Kharin, 1999; Kharin and Staleuskaya, 2000); trend and regression functions do not belong to the declared parametric family (Kharin and Staleuskaya, 2000; Kharin, 2000), etc. Unfortunately, the forecasting procedures, which are optimal under hypothetical models, often lose their optimality and become unstable under a little distorted hypothetical models. That is why the topical problem of robustness evaluation for the traditionally used forecasting procedures under distortions is considered in this paper.

Note, that the majority of publications on robustness in statistical time series analysis are concentrated on estimation of parameters and hypotheses testing. Although these problems are fundamentals, they do not overlap the problem of statistical forecasting. The distinctive feature of this paper is its orientation to the analysis of robustness of traditional forecasting procedures under model distortions by means of the mean square risk.

2 Robustness Characteristics

Let $x_t \in \mathbb{R}$, $t \in \mathbb{Z}$ be an observed time series with discrete time t, $X = (x_1, \dots, x_T)' \in \mathbb{R}^T$ be a vector of observations for T time units, T be an observation duration, $x_{T+\tau} \in \mathbb{R}$ be an unknown random variable to be forecasted, $\tau \geq 1$ be a "forecasting horizon". The probability model of the observed time series under distortions is described by a family of probability measures $\left\{P^{\varepsilon}_{T,\theta^0}(A),\ A \in \mathcal{B}^T : T \in \mathbb{N}\right\}$, where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$, $\theta^0 \in \Theta \subseteq \mathbb{R}^m$ is a true unknown value of the vector of parameters, $\varepsilon \in [0, \varepsilon_+]$ is a distortion level, and $\varepsilon_+ \geq 0$ is its admissible maximal value.

Let us define a statistical forecast $\hat{x}_{T+\tau}$ by a statistic:

$$\hat{x}_{T+\tau} = f_{T+\tau}(X). \tag{1}$$

We will evaluate performance of the forecast (1) by the functional called the point mean square risk of forecasting:

$$\rho_\varepsilon = \rho_\varepsilon\left(f_{T+\tau}; \theta^0\right) = \mathbb{E}_\varepsilon\left\{(\hat{x}_{T+\tau} - x_{T+\tau})^2\right\} \geq 0, \tag{2}$$

where $\mathbb{E}_\varepsilon\{\cdot\}$ means expectation w.r.t. the measure $P^{\varepsilon}_{T+\tau,\theta^0}\{\cdot\}$. If the distortions are absent ($\varepsilon = 0$) and we have the hypothetical model, the functional (2) will be called the hypothetical point risk: $\rho_0\left(f_{T+\tau}; \theta^0\right)$.

To eliminate dependence on the unknown parameter θ^0 in (2) we will also use the integral risk:

$$r_\varepsilon = r_\varepsilon\left(f_{T+\tau}\right) = \int_\Theta \rho_\varepsilon\left(f_{T+\tau}; \theta\right) \pi(\theta) d\theta \geq 0, \tag{3}$$

where $\pi(\theta)$ is a weight function, e.g. the prior probability density function if θ^0 is the random vector. Note, that for simple models, e.g. for the trend model (Kharin, 2000), the point risk (2) can be independent of θ^0. In this case we have $r_\varepsilon = \rho_\varepsilon$.

We define also the guaranteed (upper) risk:

$$r_+ = r_+\left(f_{T+\tau}\right) = \sup_{0 \leq \varepsilon \leq \varepsilon_+} r_\varepsilon\left(f_{T+\tau}\right), \tag{4}$$

where supremum is w.r.t. all admissible distortions of the used hypothetical model.

Let $\hat{x}^0_{T+\tau} = f^0_{T+\tau}(X; \theta^0)$ be the optimal forecast for the unknown future state $x_{T+\tau}$ under the known value θ^0 and absence of distortions, for which the risk functional $\rho_0\left(f_{T+\tau}; \theta^0\right)$ has the minimal value $\rho_0 = \rho_0\left(f^0_{T+\tau}; \theta^0\right)$. In practice, the family of the so-called "plug-in" forecasts is often used:

$$\hat{x}_{T+\tau} = f_{T+\tau}(X) := f^0_{T+\tau}(X; \hat{\theta}), \tag{5}$$

where $\hat{\theta} \in \mathbb{R}^m$ is a consistent statistical estimator of θ^0 based on the observed time series X. The forecast (5) is called the asymptotically optimal forecast if

$\lim_{T\to\infty}\left(\rho_0\left(f_{T+\tau};\theta^0\right)-\rho_0\right)=0$. The relative increment of the guaranteed risk (4) w.r.t. the hypothetical risk $\left(r_0=\int\limits_{\Theta}\rho_0\left(f^0_{T+\tau};\theta\right)\pi(\theta)d\theta>0\right)$:

$$\kappa=\kappa\left(f_{T+\tau}\right)=\left(r_+\left(f_{T+\tau}\right)-r_0\right)/r_0\geq 0 \tag{6}$$

is called the instability coefficient (Kharin, 1996, 1999). For any $\delta>0$ the value

$$\varepsilon^*=\varepsilon^*(\delta)=\sup\left\{\varepsilon:\kappa\left(f_{T+\tau}\right)\leq\delta\right\} \tag{7}$$

is called the δ-admissible distortion level (Kharin, 1996, 1999). This value is very desirable in applications. It indicates the maximal level of distortions for which the relative increment of the risk is yet not greater than the fixed value $\delta\cdot 100\%$. The smaller the value κ is and the greater the value ε^* is, the more robust the forecast $f_{T+\tau}(\cdot)$ is. The statistic $f^*_{T+\tau}(X)$ is called the minimax robust forecast if the instability coefficient (6) is minimal:

$$\kappa\left(f^*_{T+\tau}\right)=\inf_{f_{T+\tau}(\cdot)}\kappa\left(f_{T+\tau}\right). \tag{8}$$

Note, that in statistical estimation and testing hypothesis the most productive robustness concepts were introduced by the school of P. Huber (Huber, 1981; Rieder, 1994) (the minimax approach), and by the school of F. Hampel (Hampel et al., 1986) (the approach based on influence functions). For statistical forecasting problems we use here the robustness concept similar to the Γ-minimax risk criterion and the Γ-minimax regret risk criterion introduced by J. Berger (Berger, 1985).

3 Robustness in Trend Forecasting

Let the hypothetical model be a linear trend model:

$$x_t=\theta^{0'}\psi(t)+u_t,\quad t\in\mathbb{Z}, \tag{9}$$

where $\psi(t)=(\psi_j(t))\in\mathbb{R}^m$ is the vector of some m linearly independent functions, $\{u_t\}$ are i.i.d. random variables with zero mean $\mathbb{E}\{u_t\}=0$ and a bounded variance $\mathbb{D}\{u_t\}=\sigma^2<+\infty$.

According to (9), the optimal hypothetical forecast is $\hat{x}^0_{T+\tau}=\theta^{0'}\psi(T+\tau)$ and attains the minimal risk value $\rho_0=r_0=\sigma^2$. The forecast traditionally used for the model (9) is the asymptotically optimal "plug-in" forecast (5) based on the least-squares (LS)-estimator:

$$\hat{x}_{T+\tau}=\hat{\theta}'\psi(T+\tau),\quad \hat{\theta}=\left(\Psi'_T\Psi_T\right)^{-1}\Psi'_TX, \tag{10}$$

where $\Psi_T=(\psi_j(i))$ is the $(T\times m)$-matrix.

3.1 The Case of "Outliers"

Let the hypothetical model (9) be distorted by Tukey–Huber "outliers", so the observed state of the investigated process is

$$x_t = \theta^{0\prime}\psi(t) + u_t + \xi_t \cdot v_t, \quad t \in \mathbb{Z}, \tag{11}$$

where $\{\xi_t\}$ are i.i.d. Bernoulli random variables: $P\{\xi_t = 1\} = \varepsilon$, $P\{\xi_t = 0\} = 1 - \varepsilon$; $\varepsilon \in [0, \varepsilon_+]$ is the probability of "outlier" appearance, $0 \leq \varepsilon_+ < \frac{1}{2}$; $\{v_t\}$ are i.i.d. random variables, $\mathbb{E}\{v_t\} = a$, $\mathbb{D}\{v_t\} = K \cdot \sigma^2$, $K \geq 0$; $\{u_t\}$, $\{\xi_t\}$, $\{v_t\}$ are jointly independent. Note, that there are three different objectives in forecasting for the model (11): 1) to forecast the true trend value $\theta^{0\prime}\psi(T+\tau)$ (see, for example, Aivazian and Mkhitarian, 1998); 2) to forecast the true undistorted state of the investigated process $y_{T+\tau} = \theta^{0\prime}\psi(T+\tau) + u_{T+\tau}$ (see, for example, Kharin, 2000); 3) to forecast the true distorted state $x_{T+\tau}$. The risk values for these three objectives are respectively: $r^{(1)} = \mathbb{E}_\varepsilon\{(f_{T+\tau}(X) - \theta^{0\prime}\psi(T+\tau))^2\}$, $r^{(2)} = \mathbb{E}_\varepsilon\{(f_{T+\tau}(X) - (\theta^{0\prime}\psi(T+\tau) + u_{T+\tau}))^2\}$, $r^{(3)} = \mathbb{E}_\varepsilon\{(f_{T+\tau}(X) - x_{T+\tau})^2\}$. Under the declared assumptions at $a = 0$ these values are functionally dependent: $r^{(2)} = r^{(1)} + \sigma^2$, $r^{(3)} = r^{(2)} + \varepsilon K \sigma^2$. So, considering in this paper the more general third objective, we can get results for the first and the second objectives.

Introduce the notation: $\mathbf{0}_T$, $\mathbf{1}_T$ are the T-vectors all elements of which are equal to 0 and 1 respectively; $(z)_+ = \max(z, 0)$; $g(T,\tau) = (g_t(T,\tau)) = \Psi_T\left(\Psi_T'\Psi_T\right)^{-1}\psi(T+\tau) \in \mathbb{R}^T$.

Theorem 1. *Let the observed time series follow a distorted trend model according to (11), where $|\Psi_T'\Psi_T| \neq 0$, and $T \geq m$. If the traditional LS-forecast (10) for the state $x_{T+\tau}$ is used, then the guaranteed risk of forecasting is*

$$r_+(T,\tau) = \left(\sigma^2 + \varepsilon_+(a^2 + K\sigma^2)\right)\left(1 + ||g(T,\tau)||^2\right) + \\ \varepsilon_+^2 a^2 \left((1 - \mathbf{1}_T' g(T,\tau))^2 - ||g(T,\tau)||^2 - 1\right).$$

Proof. Putting (10) into (2), using (11) and properties of $\{u_t\}$, $\{\xi_t\}$, $\{v_t\}$, we have:

$$\rho_\varepsilon = \sigma^2 + \varepsilon(a^2 + K\sigma^2) + \psi'(T+\tau)\Sigma_\varepsilon\psi(T+\tau) - 2\varepsilon a B_\varepsilon'\psi(T+\tau), \tag{12}$$

where $B_\varepsilon = \mathbb{E}_\varepsilon\{\hat\theta - \theta^0\}$, $\Sigma_\varepsilon = \mathbb{E}_\varepsilon\{(\hat\theta - \theta^0)(\hat\theta - \theta^0)'\}$ are the bias and the variance matrix of the LS-estimator $\hat\theta$ respectively. From (10), (11) one can find:

$$B_\varepsilon = \varepsilon a(\Psi_T'\Psi_T)^{-1}\Psi_T'\mathbf{1}_T, \ \Sigma_\varepsilon = \sigma^2(\Psi_T'\Psi_T)^{-1} + \\ \varepsilon\left((K\sigma^2 + (1-\varepsilon)a^2)(\Psi_T'\Psi_T)^{-1} + \varepsilon a^2(\Psi_T'\Psi_T)^{-1}\Psi_T'\mathbf{1}_T\mathbf{1}_T'\Psi_T(\Psi_T'\Psi_T)^{-1}\right).$$

Putting these expressions into (12) and using the notation we get the point risk, which is independent of θ^0:

$$\rho_\varepsilon = \left(\sigma^2 + \varepsilon(a^2 + K\sigma^2)\right)\left(1 + ||g(T,\tau)||^2\right) + \\ \varepsilon^2 a^2\left((1 - \mathbf{1}_T' g(T,\tau))^2 - ||g(T,\tau)||^2 - 1\right).$$

According to (3), in this case, the integral risk is $r_\varepsilon = \rho_\varepsilon$. Differentiation of r_ε w.r.t. ε shows its monotonicity, and the maximal value of risk in (4) is at $\varepsilon = \varepsilon_+$. This fact completes the proof. □

Corollary 1. *Under "outliers in variance" ($a = 0, K > 0$) the instability coefficient (6) and the δ-admissible distortion level are:*

$$\kappa(T,\tau) = ||g(T,\tau)||^2 + \varepsilon_+ K(1 + ||g(T,\tau)||^2),$$
$$\varepsilon^*(\delta, T, \tau) = \min\left\{1/2, (\delta - ||g(T,\tau)||^2)_+ K^{-1}(1 + ||g(T,\tau)||^2)^{-1}\right\}.$$

Corollary 2. *Under "outliers in mean" ($K = 0$, $a \neq 0$) the instability coefficient (6) is:*

$$\kappa(T,\tau) = ||g(T,\tau)||^2 + \varepsilon_+ (a/\sigma)^2 (1 + ||g(T,\tau)||^2) +$$
$$\varepsilon_+^2 (a/\sigma)^2 \left((1 - \mathbf{1}'_T g(T,\tau))^2 - ||g(T,\tau)||^2 - 1\right).$$

To decrease the instability coefficient κ, the so-called local-median forecasting procedure was developed and investigated (Kharin, 2000). Let $\mathcal{T} = \{1, 2, \ldots, T\}$ be the set of time moments, and $\mathcal{T}^{(l)} = \{t_1^{(l)}, \ldots, t_n^{(l)}\} \subset \mathcal{T}$ be an l-th subset of n increasing time moments ($l = 1, \ldots, L$), L be the number of different time subsets, $1 \leq L \leq \binom{T}{n}$, n be the power of these subsets ($m \leq n \leq T$). Define the corresponding submatrices of Ψ_T and X: $X_n^{(l)} = (x_{t_i^{(l)}})$; $\Psi_n^{(l)} = \left(\psi_j\left(t_i^{(l)}\right)\right)$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. The local-median (LM) forecast is defined as the sample median of L local LS-forecasts defined according to (10):

$$\tilde{x}_{T+\tau} = \underset{1 \leq l \leq L}{\text{med}} \left\{\hat{\theta}^{(l)'} \psi(T+\tau)\right\}; \ \hat{\theta}^{(l)} = \left(\Psi_n^{(l)'} \Psi_n^{(l)}\right)^{-1} \Psi_n^{(l)'} X_n^{(l)} \qquad (13)$$

is the l-th local LS-estimate by the l-th subsample $\{x_t: t \in \mathcal{T}^{(l)}\}$. If $n = T$, then the LM-forecast (13) is equivalent to the LS-forecast (10). It is shown (Kharin, 2000) that the optimal value of the subsample size n w.r.t. the robustness criterion (8) is attained at $n^* = m$.

3.2 The Case of Functional Distortions

Let the hypothetical model (9) have functional distortions:

$$x_t = \theta^{0'} \psi(t) + u_t + \lambda(t), \quad t \in \mathbb{Z}, \qquad (14)$$

where $\lambda(\cdot)$ is an unknown functional distortion of the trend.

Theorem 2. *Let the distorted trend model (14) hold, where $|\Psi'_T \Psi_T| \neq 0$, $T \geq m$. If the LS-forecast (10) for the state $x_{T+\tau}$ is used, and the distortion function $\lambda(\cdot)$ is bounded by some known limits: $0 \leq |\lambda(t)| \leq \varepsilon(t)$, $t \in \mathbb{Z}$, then the instability coefficient (6) is*

$$\kappa = ||g(T,\tau)||^2 + \left(\varepsilon(T+\tau) + \sum_{t=1}^{T} \varepsilon(t) g_t(T,\tau)\right)^2 / \sigma^2.$$

Proof. As in the proof of Theorem 1, we have using (14):

$$\rho_\varepsilon = \sigma^2 + \lambda^2(T+\tau) + \psi'(T+\tau)\Sigma_\varepsilon\psi(T+\tau) - 2\lambda(T+\tau)B'_\varepsilon\psi(T+\tau). \quad (15)$$

By (14), (10) we find Σ_ε, B_ε; putting them into (15) and using (3), we get:

$$r_\varepsilon = \sigma^2(1 + ||g(T,\tau)||^2) + \left(\lambda(T+\tau) - \sum_{t=1}^{T} g_t(T,\tau)\lambda(t)\right)^2. \quad (16)$$

Maximizing this function w.r.t. $(T+1)$ variables $\lambda(t) \in [-\varepsilon(t), +\varepsilon(t)]$, $t = 1, 2, \ldots, T, T+\tau$ according to (4), (5), we come to the declared result. □

Corollary 3. *If the bounds are not dependent on time: $\varepsilon(t) = \varepsilon_+ \geq 0$, then the instability coefficient and the δ-admissible distortion level are:*

$$\kappa(T,\tau) = ||g(T,\tau)||^2 + (\varepsilon_+/\sigma)^2 \left(1 + \sum_{t=1}^{T} |g_t(T,\tau)|\right)^2,$$

$$\varepsilon^*(\delta, T, \tau) = \sigma\sqrt{(\delta - ||g(T,\tau)||^2)_+}\left(1 + \sum_{t=1}^{T} |g_t(T,\tau)|\right)^{-1}.$$

Theorem 3. *If the conditions of Theorem 2 are modified to the inequality $|\lambda(t)| \leq \varepsilon|\theta^{0'}\psi(t)|$, $0 \leq \varepsilon \leq \varepsilon_+$, $t \in \mathbb{Z}$, i.e. we have "relative" functional distortions, then*

$$\kappa(T,\tau) = ||g(T,\tau)||^2 + \left(\frac{\varepsilon_+}{\sigma}\right)^2 \left(\left|\theta^{0'}\psi(T+\tau)\right| + \sum_{t=1}^{T}\left|g_t(T,\tau)\theta^{0'}\psi(t)\right|\right)^2.$$

Proof. The maximum of the risk functional (16) w.r.t. $T+1$ variables: $\lambda(t) \in [-\varepsilon|\theta^{0'}\psi(t)|, \varepsilon|\theta^{0'}\psi(t)|]$, $\lambda(T+\tau) \in [-\varepsilon|\theta^{0'}\psi(T+\tau)|, \varepsilon|\theta^{0'}\psi(T+\tau)|]$, is reached at

$$\lambda^*(t) = \varepsilon|\theta^{0'}\psi(t)|\,\text{sign}(g_t),\ t = 1, \ldots, T;\ \lambda^*(T+\tau) = -\varepsilon|\theta^{0'}\psi(T+\tau)|.$$

Then, using the notation (4), (5), we get the declared result. □

Theorem 4. *If the conditions of Theorem 2 are modified to $\sum_{t=1}^{T}\lambda^2(t) + \lambda^2(T+\tau) \leq \varepsilon^2$, $0 \leq \varepsilon \leq \varepsilon_+$, i.e. we have distortions with bounded l_2-metric, then the instability coefficient and δ-admissible distortion level are:*

$$\kappa(T,\tau) = ||g(T,\tau)|| + (\varepsilon_+/\sigma)^2\left(1 + ||g(T,\tau)||^2\right),$$

$$\varepsilon^*(\delta, T, \tau) = \sigma\sqrt{(\delta - ||g(T,\tau)||^2)_+ / (1 + ||g(T,\tau)||^2)}.$$

Proof. Represent (16) in the equivalent form:

$$r_\varepsilon = \sigma^2\left(1 + ||g(T,\tau)||^2\right) + \left(\Lambda'_T G(T,\tau)\right)^2, \tag{17}$$

$$\Lambda'_T = (\lambda(1),\ldots,\lambda(T),\lambda(T+\tau)),\ G'(T,\tau) = (g_1(T,\tau),\ldots,g_T(T,\tau),-1).$$

According to (4) and the condition of the theorem, evaluation of r_+ turns into the maximization problem under the quadratic restriction:

$$Q(\Lambda_T) = \Lambda'_T(G(T,\tau)G'(T,\tau))\Lambda_T \longrightarrow \max_{\Lambda_T \in \mathbb{R}^{T+1}},\ \Lambda'_T\Lambda_T \le \varepsilon_+^2.$$

The solution of this problem is well-known (Collatz, 1963):

$$\Lambda^*_T = \varepsilon_+||G(T,\tau)||^{-1}G(T,\tau),\ Q(\Lambda^*_T) = \varepsilon_+^2||G(T,\tau)||^2.$$

Taking into account the identity: $||G(T,\tau)||^2 = 1 + ||g(T,\tau)||^2$, and using (17), (4), (5), we come to the results of the theorem. □

To "robustify" the forecast (10) under functional distortions (14), we use the special M-estimator $\tilde{\theta}$ in (10) instead of $\hat{\theta}$:

$$\tilde{\theta} = \arg\min_\theta \sum_{t=1}^{T} \rho\left(x_t - \theta'\psi(t)\right),\ \rho(z) = \begin{cases} (z+\Delta)^2, & \text{if } z < -\Delta, \\ 0, & \text{if } |z| \le \Delta, \\ (z-\Delta)^2, & \text{if } z > \Delta, \end{cases}$$

where $\Delta \ge 0$ is a parameter. Note, that this estimator turns into the LS-estimator $\hat{\theta}$ at $\Delta = 0$. For the case, where $\varepsilon_+/\sigma >> 1$, i.e. the functional distortions $\lambda(t)$ are dominant w.r.t. the random errors u_t in (14), the optimal value of Δ minimizing the instability coefficient κ was found by Monte Carlo simulations: $\Delta^* \approx 0.8\varepsilon_+$. In Figure 1 the dependence of the sample risk $\hat{r}$ on $K = \Delta/\varepsilon_+$, calculated by Monte Carlo simulations, is plotted for the special case of (14), where $T = 20$, $\tau = 1$, $m = 3$, $\psi(t) = (1,t,t^2)'$, $\theta^0 = (3,-0.5,0.05)'$, $\sigma = 0.1$, $\varepsilon_+ = 0.5$, $\lambda(t) = \varepsilon_+\cos(t)$.

4 Robustness in Parametric Regression Forecasting

Let the dynamics of an investigated stochastic system satisfy a multiple linear regression model under Tukey–Huber "outliers" and "errors-in-regressors":

$$x_t = {\theta^0}' z_t + u_t,\ \tilde{z}_t = z_t + \varepsilon_2 v_t,\ t \in \mathbb{Z}, \tag{18}$$

where $z_t \in \mathbb{R}^m$ is the non-observable nonrandom true vector of regressors (independent variables, exogenous variables), $\tilde{z}_t \in \mathbb{R}^m$ is the observable vector of regressors, $v_t \in \mathbb{R}^m$ is the sequence of i.i.d. random errors in regressors registration with the standard Gaussian distribution:

$$\mathcal{L}\{v_t\} = \mathcal{N}_m(\mathbf{0}_m, \mathbf{I}_m); \tag{19}$$

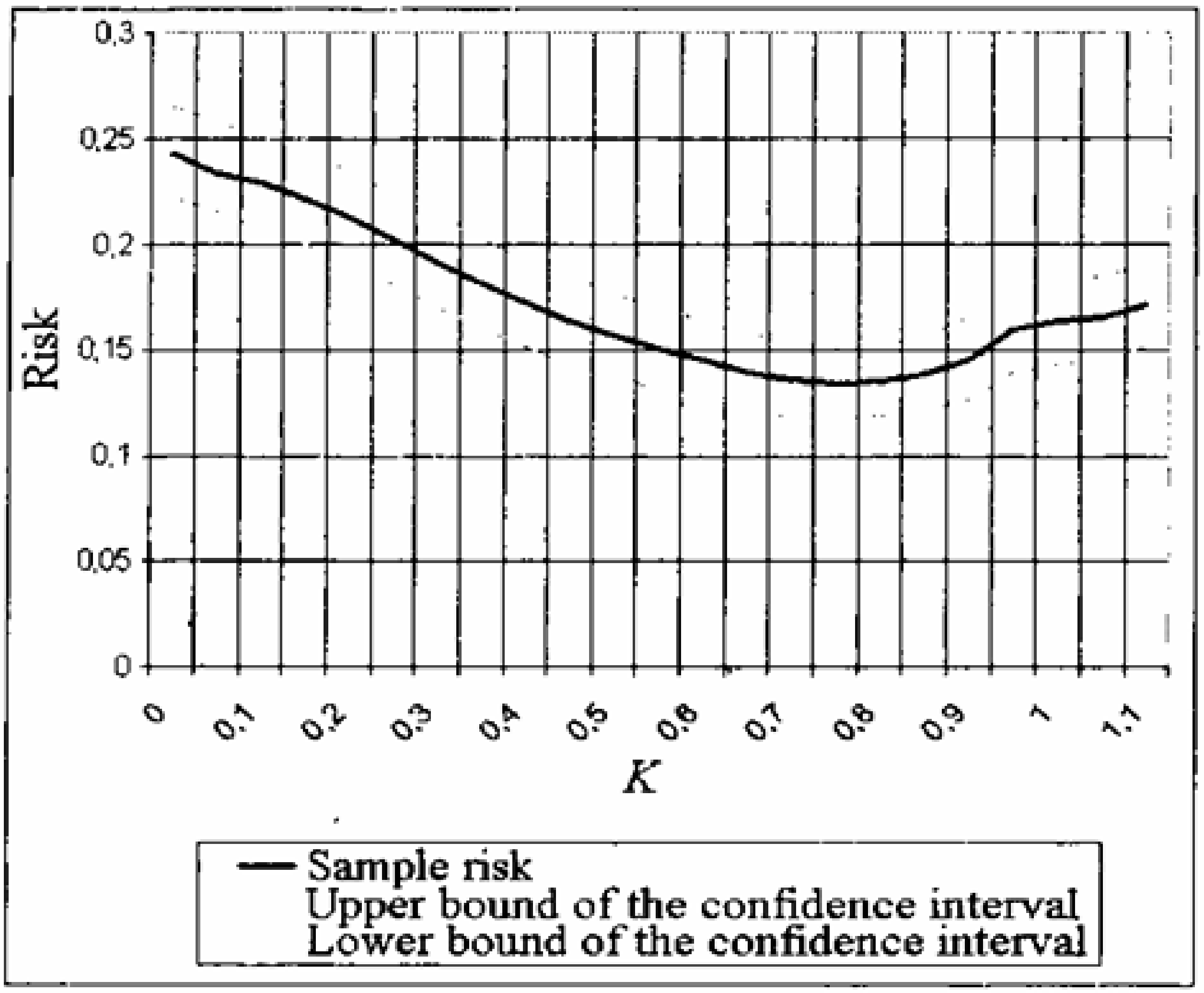

Fig. 1. Dependence of the risk on K

$\varepsilon_2 \in [0, \varepsilon_{2+}]$ is a level of distortions in regressors registration; $u_t \in \mathbb{R}$ is the sequence of independent observation errors with "outliers":

$$\mathcal{L}\{u_t\} = (1-\varepsilon_1)\mathcal{N}_1(0,\sigma^2) + \varepsilon_1 \mathcal{N}_1(0,(1+\gamma)\sigma^2), \tag{20}$$

$\gamma \geq 0$ is the coefficient indicating the increment of the variance for "outliers"; $\varepsilon_1 \in [0, \varepsilon_{1+}]$ is a level of "outliers", i.e. the expected proportion of "outliers". It is assumed that $\{u_t\}$, $\{v_t\}$ are jointly independent sequences.

The traditional forecasting procedure is defined in Anderson (1971) (if $|\tilde{Z}'\tilde{Z}| \neq 0$):

$$\hat{x}_{T+\tau} = \hat{\theta} z_{T+\tau},\ \hat{\theta} = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'X, \tag{21}$$

where $\tilde{Z}' = (\tilde{z}_1, \cdots, \tilde{z}_T)$, $Z' = (z_1, \cdots, z_T)$, $X = (x_1, \ldots, x_T)' \in \mathbb{R}^T$, $z_{T+\tau} \in \mathbb{R}^m$ is the given value of the regressor and $\hat{\theta} \in \mathbb{R}^m$ is the LS-estimator of θ^0.

Introduce the notation: $\mathcal{O}(\varepsilon)$ is the Landau symbol; $\mathcal{O}_{L_2}(\varepsilon)$ is a random vector $\eta \in \mathbb{R}^s$, for which $\mathbb{E}\{||\eta||^2\} = \mathcal{O}(\varepsilon)$;

$$\nu_k = z'_{T+\tau}(Z'Z)^{-k} z_{T+\tau} \geq 0,\ k = 1,2;\ D = \int_{\mathbb{R}^m} ||\theta||^2 \pi(\theta) d\theta \geq 0. \tag{22}$$

Theorem 5. *If the multiple linear regression model is distorted according to (18) – (20) by "outliers" and "errors-in-regressors" and $|Z'Z| \neq 0$, then the guaranteed upper risk (4) for the LS-forecast (21) satisfies the asymptotic expansion:*

$$r_+(T,\tau) = \sigma^2 + \sigma^2\nu_1 + \varepsilon_{1+}\gamma\sigma^2 + \varepsilon_{1+}\gamma\sigma^2\nu_1 + \varepsilon_{2+}^2\left(D\nu_1 + \sigma^2(1+\varepsilon_{1+}\gamma)\left((T-m)\nu_2 + \nu_1 \operatorname{tr}\left((Z'Z)^{-1}\right)\right)\right) + \mathcal{O}(\varepsilon_+^3). \tag{23}$$

Proof. From (2), (18), (21) and independence of $u_{T+\tau}$, $\hat{\theta}$, we have:

$$\rho_\varepsilon = \mathbb{E}\{u_{T+\tau}^2\} + z'_{T+\tau}\mathbb{E}\{(\hat{\theta}-\theta^0)(\hat{\theta}-\theta^0)'\}z_{T+\tau}. \tag{24}$$

Using (21), the equation: $\tilde{Z} = Z + \varepsilon_2 V$, $V = (v_1, \cdots, v_T)'$, and matrix properties from the Section 5.8 (Horn and Johnson, 1986), we get the stochastic expansion:

$$\hat{\theta} - \theta^0 = (\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'U - \varepsilon_2(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}'V\theta^0 = (Z'Z)^{-1}Z'U - \varepsilon_2\Big((Z'Z)^{-1}(V'Z + Z'V)(Z'Z)^{-1}Z'U - (Z'Z)^{-1}V'U + (Z'Z)^{-1}Z'V\theta^0\Big) + \mathcal{O}_{L_2}(\varepsilon_2^2)\mathbf{1}_m,$$

where $U = (u_1, \ldots, u_T)' \in \mathbb{R}^T$. This expansion and (19), (20) allow to get the asymptotic expansion for the variance matrix $\mathbb{E}\{(\hat{\theta}-\theta^0)(\hat{\theta}-\theta^0)'\}$. Putting it into (24), using (3), (4) and (22), we come to (23). □

In (23), one can see six main terms: the first term is the hypothetical risk $r_0 = \sigma^2$ (under full information on θ^0 without any distortions); the second term is generated by "finiteness" of observation time T; the third is generated by "outliers"; the fourth is the result of joint influence of "finiteness" and "outliers"; the fifth term $\varepsilon_{2+}^2 D\nu_1$ is the result of joint influence of "finiteness" of T and "errors-in-regressors"; the sixth term is the result of joint influence of all these three factors.

Corollary 4. *The instability coefficient (6) is*

$$\kappa(T,\tau) = \nu_1 + \varepsilon_{1+}\gamma(1+\nu_1) + \varepsilon_{2+}^2\left((D/\sigma^2)\nu_1 + (1+\varepsilon_{1+}\gamma)\left((T-m)\nu_2 + \nu_1 \operatorname{tr}((Z'Z)^{-1})\right)\right) + \mathcal{O}(\varepsilon_+^3).$$

5 Robustness in AR-forecasting

Let the hypothetical model of the observed time series be the autoregressive $AR(m)$-model:

$$x_t = \theta^{0\prime} X_{t-1} + u_t,\ t \in \mathbb{Z}, \tag{25}$$

where $X_{t-1} = (x_{t-1}, \ldots, x_{t-m})' \in \mathbb{R}^m$, $X_0 = \mathbf{0}_m$, $\{u_t\}$ are i.i.d. random variables with the Gaussian distribution $\mathcal{L}\{u_t\} = \mathcal{N}_1(0,\sigma^2)$. The optimal hypothetical forecast for (25) is defined recursively (Abraham and Ledolter, 1989):

$$\hat{x}_{T+j}^0 = \theta^{0\prime}\hat{X}_{T+j-1},\ j = 1, \ldots, \tau, \tag{26}$$

where $\hat{X}_{T+j-1} = \left(\hat{x}_{T+j-1}^0, \ldots, \hat{x}_{T+j-m}^0\right)'$, and $\hat{x}_s^0 = x_s$ for $s \le T$.

5.1 The Case of Specification Errors and Functional Distortions

Let the true value θ^0 be unknown and its approximate value $\theta \in \mathbb{R}^m$ be used in the forecasting procedure (26):

$$\hat{x}_{T+j} = \theta' \hat{X}_{T+j-1}, \quad j = 1, \ldots, \tau, \tag{27}$$

so we have a specification error $\theta - \theta^0$. Assume also that the innovation process in (25) is corrupted by functional distortions $\lambda(\cdot)$:

$$x_t = \theta^{0'} X_{t-1} + u_t + \lambda(t), \quad t \in \mathbb{Z}. \tag{28}$$

Denote the $(m \times m)$-block matrices:

$$B_0 = \begin{pmatrix} & \theta^{0'} & \\ \cdots\cdots & \cdots & \cdots\cdots \\ \mathbf{I}_{m-1} & \vdots & \mathbf{0}_{m-1} \end{pmatrix}, \quad B = \begin{pmatrix} & \theta' & \\ \cdots\cdots & \cdots & \cdots\cdots \\ \mathbf{I}_{m-1} & \vdots & \mathbf{0}_{m-1} \end{pmatrix};$$

$B_{i\cdot}$ is the i-th row of B, $B_{\cdot i}$ is the i-th column of B, $(B)_{kl}$ is the (k, l)-th element of B. Assume that all eigenvalues of B_0 are inside the unit circle.

Theorem 6. *If only the nonrandom parametric specification error $\theta - \theta^0 \in \mathbb{R}^m$ is present, then the point risk (2) for the forecast (27) is*

$$\rho_0\left(\theta^0, \theta; T, \tau\right) = \sigma^2 \left(1 + \sum_{j=1}^{\tau-1} \left(\left(B_0^j\right)_{11}\right)^2 + \sum_{t=1}^{T} \left(\left(\left(B^\tau - B_0^\tau\right) B_0^{t-1}\right)_{11}\right)^2\right).$$

Proof. According to (27), (28), the notation, and Anderson (1971), we have at $\lambda(\cdot) = 0$:

$$\begin{gathered} x_{T+\tau} = (B_0^\tau)_{1\cdot} X_T + \textstyle\sum_{i=0}^{\tau-1} (B_0^i)_{11} u_{T+\tau-i}, \; \hat{x}_{T+\tau} = (B^\tau)_{1\cdot} X_T; \\ X_t = \textstyle\sum_{i=0}^{t-1} B_0^i U_{t-i}, \; U_t = (u_t, u_{t-1}, \ldots, u_{t-m+1})' \in \mathbb{R}^m. \end{gathered} \tag{29}$$

Using the properties of $\{u_t\}$ we get by (2)

$$\rho_0 = \sigma^2 \left(1 + \sum_{i=1}^{\tau-1} \left((B_0^i)_{11}\right)^2\right) + (B^\tau - B_0^\tau)_{1\cdot} \Sigma_T (B^\tau - B_0^\tau)'_{1\cdot}, \tag{30}$$

where according to (29) $\Sigma_T = \mathbb{E}\{X_T X_T'\} = \sigma^2 \sum_{i=0}^{T-1} B_0^{2i}$. Putting this into (30) and making equivalent transformations we come to the result. □

Theorem 7. *If the functional distortions in (28) are bounded:*

$$T^{-1} \sum_{t=1}^{T} \lambda^2(t) \le \varepsilon_{(1)}^2, \quad \tau^{-1} \sum_{t=T+1}^{T+\tau} \lambda^2(t) \le \varepsilon_{(2)}^2,$$

where $\varepsilon_{(1)}, \varepsilon_{(2)} \geq 0$ are distortion levels for the "observation interval" and for the "forecasting interval" respectively, then the guaranteed point risk for the forecast (27) is

$$\rho_+ \left(\theta^0, \theta; T, \tau\right) = \rho_0 \left(\theta^0, \theta; T, \tau\right) +$$

$$+ \left(\varepsilon_{(1)} \sqrt{T \sum_{i=0}^{T-1} \left(\left(\left(B^\tau - B_0^\tau\right) B_0^i \right)_{11} \right)^2} + \varepsilon_{(2)} \sqrt{\tau \sum_{i=0}^{\tau-1} \left(\left(B_0^i \right)_{11} \right)^2} \right)^2 .$$

Proof. As in Theorem 6, using (27), (28), we get the point risk under functional distortions:

$$\rho_\varepsilon = \rho_0(\theta^0, \theta; T, \tau) + (\varepsilon_{(1)} q_T' \mathbf{e}_T + \varepsilon_{(2)} \bar{q}_\tau' \bar{\mathbf{e}}_\tau)^2, \tag{31}$$

where $q_T = \left(((B^\tau - B_0^\tau) B_0^0)_{11}, \ldots, ((B^\tau - B_0^\tau) B_0^{T-1})_{11} \right)' \in \mathbb{R}^T$,

$$\bar{q}_\tau = -\left((B_0^0)_{11}, \ldots, (B_0^{\tau-1})_{11} \right)' \in \mathbb{R}^\tau, \ \bar{\mathbf{e}}_T = (e_{T+\tau}, \ldots, e_{T+1})' \in \mathbb{R}^\tau,$$

$$\mathbf{e}_T = (e_T, \ldots, e_1)' \in \mathbb{R}^T, e_t = \{ \frac{\lambda(t)}{\varepsilon_{(1)}}, 1 \leq t \leq T; \frac{\lambda(t)}{\varepsilon_{(2)}}, T < t \leq T + \tau \}.$$

Maximizing ρ_ε as a quadratic functional w.r.t. $T+\tau$ variables $\mathbf{e}_T, \bar{\mathbf{e}}_\tau$ under quadratic restrictions derived by the conditions of the theorem: $\mathbf{e}_T' \mathbf{e}_T \leq T$, $\bar{\mathbf{e}}_\tau' \bar{\mathbf{e}}_\tau \leq \tau$, we get the statement of the theorem. □

A Monte Carlo study has shown good fitting of theoretical and simulation results. In Figure 2 the solid line plots dependence of the theoretical risk (31) on ε, the dashed line – the sample risk calculated by $N = 10000$ Monte Carlo simulations, and the dotted line – the minimal risk $\rho_0 = \sigma^2$ for the special case of the model (28), where $T = 40$, $\tau = 2$, $m = 2$, $\theta^0 = (0.3, 0.4)'$, $\theta = (0.4, 0.5)'$, $\sigma = 1$, $\lambda(t) = \varepsilon \sin(t)$.

5.2 LS-forecasting under Functional Distortions

The LS-forecast is defined by (27), where $\theta = \hat{\theta}$ is the LS-estimator:

$$\hat{\theta} = \left(\sum\nolimits_{t=1}^{T} X_{t-1} X_{t-1}' \right)^{-1} \sum\nolimits_{t=1}^{T} x_t X_{t-1}. \tag{32}$$

Because of nonlinear dependence of the forecast $\hat{x}_{T+\tau}$ on the time series X and of dependence between $\hat{\theta}$ and X_T (see (29) and (27)) it is difficult to present simple formulas for the robustness characteristics. For simplicity, consider here the special situation only, where the estimate $\hat{\theta}$ is found according to (32) by the time series $\tilde{X} = (\tilde{x}_1, \ldots, \tilde{x}_{T_0})' \in \mathbb{R}^{T_0}$, which is independent of X and is free of distortions. Such a situation takes place when "the process of training of the forecasting algorithm" (27) by the data $\tilde{X}$ is separated from "the process of forecasting" on the data X (Aivazian and Mkhitarian, 1998).

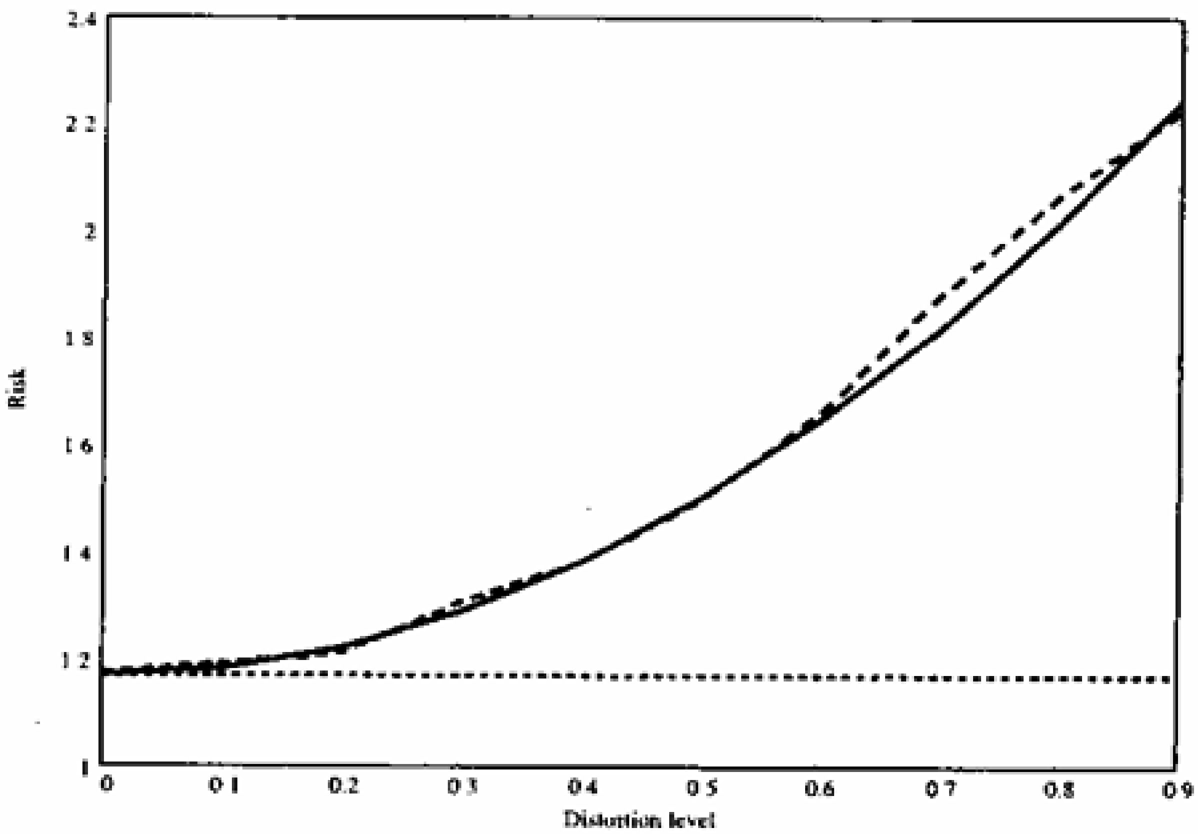

Fig. 2. Dependence of the risk on the distortion level ε

Theorem 8. *Let the AR(m)-model (28) hold, where* $T^{-1}\sum_{t=1}^{T}\lambda^2(t) \le \varepsilon_+^2$; $\lambda(t) = 0$, $t > T$. *Suppose the forecast at* $\tau = 1$ *is obtained by the algorithm (27), where* $\hat{\theta}$ *is the LS-estimator (32) based on a time series* $\tilde{X} \in \mathbb{R}^{T_0}$ *of the length* T_0, *which is independent of* X *and is free of distortions. If the variance matrix of* $\hat{\theta}$ *is* $\Sigma = \mathbb{E}\{(\hat{\theta}-\theta^0)(\hat{\theta}-\theta^0)'\} = (T_0)^{-1}F$, *then the guaranteed point risk is*

$$\rho_+(\theta^0; T, T_0) = \sigma^2 + \frac{\sigma^2}{T_0}\sum_{t=1}^{T}(B_0^{t-1})'_{\cdot 1}F(B_0^{t-1})_{\cdot 1} + \varepsilon_+^2\frac{T}{T_0}\mu_{\max}(G_T),$$

where $\mu_{\max}(G_T)$ *denotes the maximal eigenvalue of* $(T \times T)$*-matrix*

$$G_T = \left((B_0^{i-1})'_{\cdot 1}F(B_0^{j-1})_{\cdot 1}\right), \quad i,j = 1,\dots,T.$$

Proof. By (31) and Theorem 7 for $\varepsilon_{(1)} = \varepsilon$, $\varepsilon_{(2)} = 0$, $\tau = 1$ and fixed $\hat{\theta} = \theta^0 + \Delta$: $\rho_\varepsilon = \sigma^2 + \sigma^2\sum_{t=1}^{T}\left(\Delta'(B_0^{t-1})_{\cdot 1}\right)^2 + \varepsilon_+^2\left(\sum_{t=1}^{T}\Delta'(B_0^{t-1})_{\cdot 1}e_{T-t+1}\right)^2$. Applying expectation w.r.t. $\Delta = \hat{\theta} - \theta^0$ under the conditions of the theorem and maximizing w.r.t. $(e_1,\dots,e_T)$ under the quadratic restriction: $\sum_{t=1}^{T}e_t^2 \le T$, we get the declared result. □

Note also, that in Fursa (1996) the robustness characteristics (4), (6), (7) were evaluated for the case of additive "outliers" in the AR(m) model.

6 Conclusion

The results presented here provide a statistician with quantitative estimates of the guaranteed upper risk (4), the instability coefficient (6) and the δ-admissible distortion level (7) in "plug-in" statistical forecasting of time series under some typical distortions of the underlying hypothetical models, i.e. trend, regression and autoregressive models. These estimates reveal the influence of distortions on the risk, indicate the limits of the "safe" forecasting by traditional algorithms under distortions, and outline approaches to the "robustification" of forecasting procedures.

The theoretical results were tested by Monte Carlo modelling and applied in special software packages: ROSTAN (RObust STatistical ANalysis); STATFOR (STATistical FORecasting) (Kharin et al., 2000); STAT-RO (STATistics-RObust); SEMF (System of Econometric Modelling and Forecasting) (Kharin and Staleuskaya, 2000).

Acknowledgements

The author gratefully acknowledge the careful reading and very pertinent remarks by two anonymous referees, which lead to a substantial improvement of the initial version of this paper.

This research was supported by the International Science and Technology Center, project B-705.

References

B. Abraham and J. Ledolter. *Statistical methods for forecasting*. Wiley, New York, 1989.

S. Aivazian and V. Mkhitarian. *Applied statistics and essentials of econometrics*. Unity, Moscow, 1998.

T.W. Anderson. *The statistical analysis of time series*. Wiley, New York, 1971.

J.O. Berger. *Statistical decision theory and Bayesian analysis*. Springer, New York, 1985.

M. Boutahar and C. Deniau. A proof of asymptotic normality for some VARX Models. *Metrika*, 42:331–339, 1995.

B.L. Bowerman. *Forecasting and time series: an applied approach*. Duxbury Press, Belmont, MA, 1993.

L. Collatz. *Eigenwertaufgaben mit technischen Anwendungen*. Akad. Verlag, Leipzig, 1963.

R. Dahlhaus and W. Wefelmeyer. Asymptotically optimal estimation in misspecified time series models. *Ann. Statist.*, 24:952–974, 1996.

R. Fursa. On robustness of forecasting of AR(p) time series. In I.V. Gaishun, editor, *Proceedings of VII Belarus Math. Conf. BSU*, pages 63–67, Minsk, 1996.

J.D. Hamilton. *Time series analysis*. Princeton Univ. Press, 1994.

F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust statistics: The approach based on influence functions*. Wiley, New York, 1986.

R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge Univ. Press, 1986.

P.J. Huber. *Robust statistics*. Wiley, New York, 1981.

Y. Kharin. *Robustness in statistical pattern recognition*. Kluwer Academic Publishers, Dordrecht / Boston / London, 1996.

Y. Kharin. Robustness of statistical classification under nonparametric distorsions. *Math. Met. of Statist.*, 8:84–98, 1999.

Y. Kharin. Robust forecasting of a parametric trend under "ouliers". *Studies in Classif. and Data Anal.*, 17:197–206, 2000.

Y. Kharin and S. Staleuskaya. Robustification of "approximating approach". In T. Kollo, editor, *New Trends in Prob. and Statist. TEV*, pages 143–150, Vilnius, 2000.

Y. Kharin, M. Abramovich, S. Agievich, and A. Akinfin. *User's guide on STATFOR software*. BSU, Minsk, 2000.

D. Peña, G. Tiao, and R. Tsay. *A course in time series analysis*. Wiley, New York, 2001.

H. Rieder. *Robust asymptotic statistics*. Springer, New York, 1994.